# Learning non-Markovian physics from data ☆

David González [a], Francisco Chinesta [b], Elías Cueto [a,*]

[a] Aragon Institute of Engineering Research (I3A), Universidad de Zaragoza, Maria de Luna 3, E-50018 Zaragoza, Spain
[b] ESI Chair and PIMM Lab, ENSAM ParisTech, 155 Boulevard de l'Hôpital, 75013 Paris, France

*Keywords:*
Machine learning
Generalized Langevin Equation
Mori-Zwanzig projection
History-dependent physics
GENERIC

A B S T R A C T

We present a method for the data-driven learning of physical phenomena whose evolution in time depends on history terms. It is well known that a Mori-Zwanzig-type projection produces a description of the physical phenomena that depends on history, and also incorporates noise. If the data stream is sampled from the projected Mori-Zwanzig manifold, the description of the phenomenon will always depend on one or more unresolved variables, *a priori* unknown, and will also incorporate noise.

The present work introduces a novel technique able to unveil the presence of such internal variables—although without giving it a precise physical meaning—and to minimize the inherent noise. The method is based upon a refinement of the scale at which the phenomenon is described by means of *kernel*-PCA techniques. By learning the metriplectic form of the evolution of the physics, the resulting approximation satisfies basic thermodynamic principles such as energy conservation and positive entropy production.

Examples are provided that show the potential of the method in both discrete and continuum mechanics.

## 1. Introduction

Computational mechanics, as any other scientific discipline, could get benefit from the widespread availability of data and the ease of transmission through mobile networks. The different forms of machine learning allow, in turn, to obtain knowledge from data. It is not uncommon, nowadays, that even scientific principles can be unveiled in closed form from data without any human intervention [1,2].

In general, we are not even interested, when learning physical laws, in obtaining closed form expressions. For instance, in Kirchdoerfer and Ortiz [3] or Ayensa-Jimenez et al., [4] raw data are employed to substitute constitutive equations in solid mechanics. In these works, equilibrium and compatibility equations are considered in a traditional way, while constitutive laws are substituted by the closest experimental strain-stress pair satisfying equilibrium and compatibility equations. Many other references have tackled this promising way of employing data recently [5–8]. New concepts as the *constitutive manifold* have arisen [9–11] Other approaches employ neural networks, for instance [12–14].

A careful analysis of these different approaches reveals, however, some problems not yet fully solved nor understood. One is the thermodynamic admissibility of the resulting approximations to the physical phenomenon. Of course, the un-

☆ This project has been partially funded by the ESI Group through the ESI Chair at ENSAM ParisTech and through the project "Simulated Reality" at the University of Zaragoza. The support of the Spanish Ministry of Economy and Competitiveness through grant number CICYT-DPI2017-85139-C2-1-R and by the Regional Government of Aragon, through the project T24_20R, and the European Social Fund, are also gratefully acknowledged.

* Corresponding author.
*E-mail addresses:* gonzal@unizar.es (D. González), francisco.chinesta@ensam.eu (F. Chinesta), ecueto@unizar.es (E. Cueto).

veiled structure of these approaches must fulfill basic thermodynamic principles such as conservation of energy and entropy production for isolated systems. Recently, the authors have suggested the employ of these data-driven approaches so as to obtain, by regression, the GENERIC structure of the phenomenon at hand [15,16]. GENERIC (an acronym for General Equation for Non-Equilibrium Reversible-Irreversible Coupling [17,18]) is a completely general equation for the time evolution of a physical system out of equilibrium. By obtaining the value of the terms in the GENERIC equation from data, one can ensure the fulfillment of the laws of thermodynamics. In addition, this approach gives rise to a thermodynamic interpretation of machine learning of physical laws that gives insight to the problem and greatly contributes to its understanding. This interpretation has been developed by the authors recently [19] and will be reviewed and adapted to the objectives of this paper in Section 2 for completeness.

Another big problem in the development of data-driven approaches to continuum mechanics lies in the presence of history-dependent phenomena. Although some approaches to the problem exist [20] these are not completely general nor satisfactory, since they involve the treatment of data in the form of time series—trajectories in phase space.

A long tradition exists in the development of Generalized Langevin Equations (GLE) of phenomena for which micro-scale (usually molecular dynamics) descriptions exist [21–25]. GLE appear after projections of fully resolved data onto scales for which some details are skipped. This process is usually known in the literature as a Mori-Zwanzig projection [26,27].

Essentially, what Mori and Zwanzig said about coarse graining of physical models is that the elimination of some degrees of freedom from the fully-resolved description leads to an equation, the GLE, that includes an integral term depending on time (history) and a noise term. The interested reader can consult, among other references, the recent work by Venturi and collaborators on the data-driven approximation of the MZ equation, and particularly, of its memory term [28–32]. In general, in the above-cited references, starting from the scale of molecular dynamics, where Newton laws apply, some form of coarse graining is sought, by determining the form of the history kernel in the GLE so as to avoid its integration in time.

Here, on the contrary, we assume that experiments are performed at a scale where some form of history dependence is already present. Consequently, also noise will be present in one form or another. These appear as a consequence of eliminating relevant degrees of freedom in the phase space of the problem. In continuum mechanics this problem has traditionally been solved by the phenomenological establishment of *internal variables*. These inform us about the unresolved degrees of freedom, as well as the history dependence of the problem at hand.

However, in a purely data-driven and machine learning (thus, completely unsupervised) approach, it is not evident how to determine which is the best form for these internal variables, what is the true influence of history in the problem and the effect of the inherent noise in the experiments. To give insight about this problem, we rely on the thermodynamic interpretation of the process of machine learning of physical phenomena established by the authors [19]. This makes use, in turn, of the dynamical systems equivalence of machine learning, a parallelism already presents in some previous works [33,34].

This dynamical systems equivalence will be briefly revisited in Section 2, along with the GENERIC approach to the problem, which makes extensive use of this equivalence. This will allow us to describe the problem in thermodynamic and geometric terms. In Section 3 a toy problem will be employed to describe the approach here developed. This problem, however, does not admit a GENERIC description, since it does not represent any physical process. The full methodology will be described in Section 4.1. Section 4.2 addresses this problem for the case of viscoelastic fluids, by taking the Oldroyd-B model as a paradigm. The paper is completed with the usual discussion in Section 5.

## 2. A dynamical systems approach to thermodynamically sound machine learning

Some authors have employed a mathematical equivalence of machine learning to dynamical systems [33,34]. Assume that a set of measurements have been performed on a group of variables at discrete time instants $t$. These are stored in a vector $\boldsymbol{\xi}(t) \in \mathbb{R}^d$. Of course, we are interested in predicting their time evolution,

$$\dot{\boldsymbol{\xi}} = f(t, \boldsymbol{\xi}), \quad \boldsymbol{\xi}(0) = \boldsymbol{\xi}_0.$$

Given some time horizon $T$, we are interested in determining the flow map

$$\boldsymbol{\xi}_0 \to \boldsymbol{\xi}(T, \boldsymbol{\xi}_0).$$

Under the prism of the so-called *dynamical systems equivalence*, machine learning of the physics giving rise to the experimental measurements $\boldsymbol{\xi}(t, \boldsymbol{\xi}_0)$ is equivalent to obtaining, by regression, the expression of the function $f$. This can be done by employing neural networks or by more traditional technologies for the solution of inverse problems.

What is interesting about this way of seeing machine learning is the possibility of imposing certain properties to $f$. For instance, if for any reason we know in advance that the physics under scrutiny conserves energy, we can impose a Hamiltonian structure to the evolution dictated by $f$.

In a general framework, however, we are not allowed to impose *a priori* any conservative structure to the system. Instead, the authors have proposed recently [15,16] the employ of a more general description, coined as General Equation for Non-Equilibrium Reversible-Irreversible Coupling (GENERIC) [17,18]. Under this prism, the evolution of the system will take the form

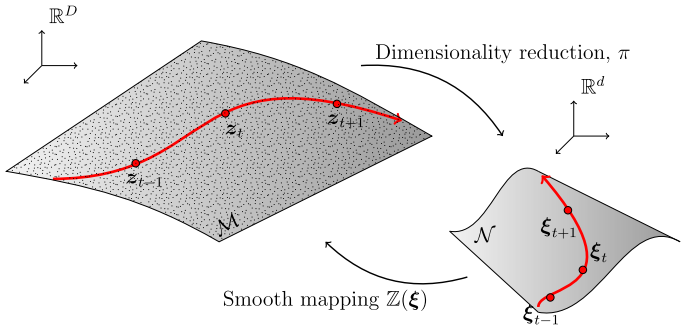



**Fig. 1.** Mori-Zwanzig projections give rise to a sequence of projections onto coarse-grained descriptions of the system, i.e., from any detailed description whose manifold structure is here represented by $\mathcal{M}$, to a coarse-grained one, here represented by its manifold structure $\mathcal{N}$. For each one of these descriptions, a particular GENERIC structure emerges, with distinct energy and entropy potentials.

$$\dot{\boldsymbol{\xi}}_t = \boldsymbol{L}(\boldsymbol{\xi}_t)\nabla E(\boldsymbol{\xi}_t) + \boldsymbol{M}(\boldsymbol{\xi}_t)\nabla S(\boldsymbol{\xi}_t), \;\; \boldsymbol{\xi}(0) = \boldsymbol{\xi}_0, \tag{1}$$

with $\boldsymbol{L}$ the so-called Poisson matrix, responsible for the reversible (Hamiltonian) part of the evolution of the system. $E$ represents the energy of the system, as a function of its particular state at time $t$, $\boldsymbol{\xi}_t$. Dissipation is introduced here by means of a second potential, entropy $S$. $\boldsymbol{M}$ represents the friction matrix, responsible for the irreversible part of the evolution of the system.

For this equation to represent valid physics, it must be supplemented by the so-called degeneracy conditions:

$$\boldsymbol{L}(\boldsymbol{\xi}) \cdot \nabla S(\boldsymbol{\xi}) = \boldsymbol{0}, \tag{2a}$$

$$\boldsymbol{M}(\boldsymbol{\xi}) \cdot \nabla E(\boldsymbol{\xi}) = \boldsymbol{0}. \tag{2b}$$

These ensure the correct fulfillment of the first and second principles of thermodynamics, i.e., by the usual choice of $\boldsymbol{L}$ to be skew-symmetric and $\boldsymbol{M}$ to be symmetric, semi-positive definite, we arrive at

$$\dot{E}(\boldsymbol{\xi}) = \nabla E(\boldsymbol{\xi}) \cdot \dot{\boldsymbol{\xi}} = \nabla E(\boldsymbol{\xi}) \cdot \boldsymbol{L}(\boldsymbol{\xi})\nabla E(\boldsymbol{\xi}) + \nabla E(\boldsymbol{\xi}) \cdot \boldsymbol{M}(\boldsymbol{\xi})\nabla S(\boldsymbol{\xi}) = 0.$$

In other words, one ensures the conservation of energy in closed systems. In the same spirit,

$$\dot{S}(\boldsymbol{\xi}) = \nabla s(\boldsymbol{\xi}) \cdot \dot{\boldsymbol{\xi}} = \nabla S(\boldsymbol{\xi}) \cdot \boldsymbol{L}(\boldsymbol{\xi})\nabla E(\boldsymbol{\xi}) + \nabla S(\boldsymbol{\xi}) \cdot \boldsymbol{M}(\boldsymbol{\xi})\nabla S(\boldsymbol{\xi}) \geq 0,$$

ensures the fulfillment of the second principle of thermodynamics.

Our approach to obtain thermodynamically consistent machine learning procedures thus consists in solving the following regression (constrained to the degeneracy conditions) within a time interval $T$:

$$\boldsymbol{\mu}^* = \{\mathsf{L}, \mathsf{M}, \mathsf{DE}, \mathsf{DS}\} = \arg\min_{\boldsymbol{\mu}} ||\boldsymbol{\xi}(\boldsymbol{\mu}) - \boldsymbol{\xi}^{\mathtt{meas}}||, \tag{3}$$

with $\boldsymbol{\xi}^{\mathtt{meas}} \subseteq \boldsymbol{\Xi}$, a subset of the total available experimental results. Here, $\boldsymbol{\mu} = \{\mathsf{L}, \mathsf{M}, \mathsf{DE}, \mathsf{DS}\}$ represent the sought parameters of the GENERIC model, the numerical approximations to the Poisson matrix $\boldsymbol{L}$, the friction matrix $\boldsymbol{M}$, the gradient of energy $\nabla E$ and the gradient of entropy, $\nabla S$. These are expressed in a piece-wise linear, finite element sense. The interested reader can consult our previous works on the topic for more details [15,16].

It is worth noting that both potentials, $E$ and $S$ depend of the chosen level of description for the system at hand. If, for instance, one chooses to obtain data from molecular dynamics simulations, where only Newtonian mechanics apply and there is no dissipation, no entropy potential will be necessary. This constitutes the most detailed level of description one could think of. However, it must be readily coarse-grained, given the immense amount of data it encompasses (position and momentum of every molecule at each time step).

Eliminating some of these degrees of freedom and retaining others is done by a Mori-Zwanzig projection, $\pi$. This results in the incorporation of noise into the less-detailed description. And fluctuation is well known to be equivalent to dissipation by the Nyquist theorem [35], thus justifying the need for the entropy potential. A graphical interpretation of this is depicted in Fig. 1. At the end of this sequence of coarse-grained descriptions lies thermodynamics, where only invariants are employed to describe the system and thus no evolution equation exists. A related approach can be found in the work by Wan et al. [36].

The machine learning procedure is seen under this prism as a dissipative process in which redundant data (think of position and momentum of the molecules) is projected onto a coarse-grained manifold in which only relevant information is manipulated [19]. But this projection $\pi$ is in general not bijective. In other words, multiple micro states give rise to the same macro state. This is finally translated into fluctuation of the results and, finally, into dissipation [37].

In the next Section we give further insight on the origin of the problem and pave the way towards its solution.

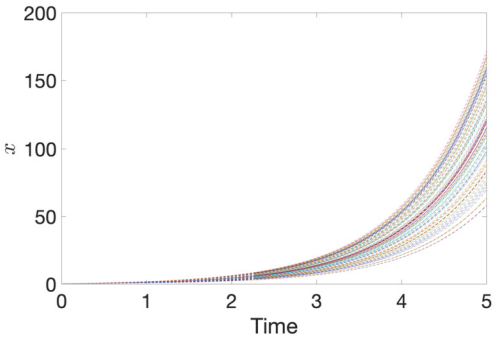



**Fig. 2.** Sampling of the time evolution of $x$. At each one of the one hundred samplings, a different $y_0$ is chosen—assuming a normal distribution—, thus giving the different trajectories. In the absence of any knowledge about the existence of $y$, this produces the apparent noise in the Mori-Zwanzig formalism.

## 3. How to learn internal variables from data

In this section we develop the basics of the proposed approach. We begin by a succinct description of the effect of unresolved variables in the final description of the system. We then briefly describe kernel-PCA methods, that play an essential role in the technique here developed.

### 3.1. An introductory example

We borrow a toy example first introduced by W. E [38]. Consider a system whose dynamics is described by two degrees of freedom,

$$\dot{x} = A_{11}x + A_{12}y, \tag{4a}$$

$$\dot{y} = A_{21}x + A_{22}y, \tag{4b}$$

with $x(0) = x_0$ and $y(0) = y_0$. If we substitute $y$ into Eq. (4a), we obtain

$$\dot{x} = \underbrace{A_{11}x}_{\text{Markovian}} \underbrace{+ A_{12}\int_0^t e^{A_{22}(t-s)}A_{21}x(s)ds}_{\text{Non-Markovian}} + \underbrace{A_{12}e^{A_{22}t}y_0}_{\text{Noise}}, \tag{5}$$

viz., a Markovian term, a non-Markovian one (thus depending on history) and a third term that depends on $y_0$ that can be interpreted as noise. This is the Mori-Zwanzig formalism for a linear case [26,27].

The phase space in which we parameterize the state of the system by pairs $\boldsymbol{z} = (x, y)$ is represented in Fig. 1 by the manifold $\mathcal{M}$. For this simple example, $D = 2$. Imagine that, for any reason—such as technical limitations or lack of knowledge—, we are allowed to sample only $x$-values, and no information from $y$ is available, even its mere existence. In that case, we look for a description of the system on a reduced-order manifold akin to $\mathcal{N} \subset \mathbb{R}^{d=1}$, in which there is only one degree of freedom instead of two: $\xi = x$.

Performing measurements on $\mathcal{N}$ will provide a set of trajectories for $x$. Here, we assumed that $x_0 = 0.5$, $y_0$ is uniformly distributed along $[-1.2, 0.8]$ and that $A_{11} = 1.0$, $A_{12} = -0.3$, $A_{21} = -0.3$, $A_{22} = 0.0$. By performing 100 different measurements, we obtain 100 slightly different curves, represented in Fig. 2. Remember that we assumed that we are not allowed to perform measurements on $y$.

The problem is thus established in the following terms: *is it possible to project back these trajectories to a manifold $\mathcal{M}^* \subset \mathbb{R}^{D^*>d}$, such that the presence of an internal variable is detected?* We do not aspire to project these experimental results back to $\mathcal{M}$, where $D^* = D$, since we are not aware of the existence of $y$. For the objectives of this work, it will be enough to establish one or more internal variables that allow us to get rid of the dependence on history and noise.

To achieve such a projection $\mathbb{Z}$, see Fig. 2, we employ kernel-Principal Component Analysis ($k$-PCA) techniques [39,40]. These are reviewed next for completeness. The reason for this choice is that $k$-PCA, a technique for manifold learning, unveils the manifold structure of data by first projecting them to a high-dimensional space, where everything is linearly separable. This property will be leveraged not to reduced order modeling of our data, but to project them back to higher dimensional manifolds $\mathcal{M}$.

### 3.2. A brief review of k-PCA

Despite of its name, $k$-PCA works, like multi-dimensional scaling, with the matrix of pairwise scalar products, the Gram matrix $\boldsymbol{S} = \boldsymbol{\Xi}^T\boldsymbol{\Xi}$, where $\boldsymbol{\Xi}$, we recall, is the matrix of snapshots obtained by sampling $\mathcal{N}$ [41]. The basic ingredient of

$k$-PCA methods, however, is appealing: data not linearly separable in $d$ dimensions, could be linearly separated if previously projected to a space in $D > d$ dimensions. In a surprising strategy, $k$-PCA projects data to an even higher dimensional space. To this end, a mapping

$$\phi : \mathcal{N} \subset \mathbb{R}^d \to \mathbb{R}^R,\ \boldsymbol{\xi} \to \boldsymbol{z} = \phi(\boldsymbol{\xi}),$$

is employed, where $R$ may be any dimension (even infinite). The true advantage of $k$-PCA is that, in practice, there is no need to explicitly determine the analytical expression of the mapping $\phi$.

To do so, we need to decompose $\boldsymbol{\Phi} = \boldsymbol{Z}^T\boldsymbol{Z}$, a matrix defined in $\mathbb{R}^{M\times M}$, into eigenvalues and eigenvectors. $M$ represents here the number of stored snapshots $\boldsymbol{z}_i$, stored in a matrix $\boldsymbol{Z}$. Previously, the mapped data $\boldsymbol{z}_i$ involved in $\boldsymbol{\Phi}$ is centered. This centering process would be impossible without knowing the mapping $\phi$. However, it can be achieved in an implicit way through a double centering.

The mean of the $j$-th column of $\boldsymbol{\Phi}$ is denoted by $\mu_i(\boldsymbol{z}_i \cdot \boldsymbol{z}_j)$, and the mean of its $i$-th row, in turn, $\mu_j(\boldsymbol{z}_i \cdot \boldsymbol{z}_j)$. The mean of all entries of $\boldsymbol{\Phi}$ will be therefore $\mu_{i,j}(\boldsymbol{z}_i \cdot \boldsymbol{z}_j)$. The double centering takes the form

$$\boldsymbol{z}_i \cdot \boldsymbol{z}_j - \mu_i(\boldsymbol{z}_i \cdot \boldsymbol{z}_j) - \mu_j(\boldsymbol{z}_i \cdot \boldsymbol{z}_j) + \mu_{i,j}(\boldsymbol{z}_i \cdot \boldsymbol{z}_j).$$

The spectral decomposition can be performed on the double centered matrix, by computing

$$\boldsymbol{\Phi} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T,$$

thus giving the reduced matrix

$$\boldsymbol{\Upsilon} = \boldsymbol{I}_{D\times M}\boldsymbol{\Lambda}^{1/2}\boldsymbol{U}^T.$$

It is worth noting that, in contrast with multi-dimensional scaling, the maximum number of strictly positive eigenvalues is not bounded by $\min(M, d)$, but by $\min(M, R)$. This opens the possibility of using $k$-PCA to map data to a higher-dimensional space, instead to a lower one, since $R$ is usually very big, even infinity.

In practice, the mapping $\phi$ is used exclusively to compute scalar products. If the images of the mapping belong to a high dimensional space, these products will involve an enormous number of multiplications. To avoid all these products, an even to compute the mapping $\phi$, a kernel function $\kappa$ is employed instead, such that it gives the value of the scalar product by $\kappa(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j) = \boldsymbol{z}_i \cdot \boldsymbol{z}_j$. This is possible thanks to Mercer's theorem [39]. It establishes that, if $\kappa(\boldsymbol{\xi}, \boldsymbol{\eta})$ is continuous, symmetric and positive definite, then it defines an inner-product in the mapped space. This property is known as the "kernel trick" and is the principal ingredient of $k$-PCA that makes it so interesting.

Among the different kernels that fulfill Mercer's condition, we have

- Polynomial kernels: $\kappa(\boldsymbol{\xi}, \boldsymbol{\eta}) = (\boldsymbol{\xi} \cdot \boldsymbol{\eta} + 1)^p$, with $p$ an arbitrary integer;
- Gaussian kernels: $\kappa(\boldsymbol{\xi}, \boldsymbol{\eta}) = \exp\left(-\frac{\|\boldsymbol{\xi}-\boldsymbol{\eta}\|^2}{2\sigma^2}\right)$ for a real $\sigma$;
- Sigmoid kernels: $\kappa(\boldsymbol{\xi}, \boldsymbol{\eta}) = \tanh(\boldsymbol{\xi} \cdot \boldsymbol{\eta} + b)$ for a real $b$.

There is no particular reason to choose one kernel or another, but their performance in the separation process. The final goal is that standard PCA performed on the mapped data is able to efficiently separate the data.

This is the key ingredient of our technique: to map coarse-grained, experimental results onto a high-dimensional manifold in dimension $R \gg d$ and then map back, via PCA, to a manifold $\mathcal{M}^* \subset \mathbb{R}^{D^*}$, where $D^* \geq D$, the exact manifold where the intrinsic variables are described.

The resulting procedure will look like a mapping from $\mathcal{N}$ to $\mathcal{M}$, or at least, to some approximation of it. We review its performance on the toy example just introduced below.

### 3.3. Finding the intrinsic variable

Of course, we know that the different curves in Fig. 2 are due to the existence of $y$, the intrinsic variable, and its different initial values $y_0$, different for each sampling. However, in the absence of any information on $y$, the proposed method should be able to first unveil the existence of a hidden variable, and to equip the experimental results with some topology, dictated by $y$, so that we can interpolate among experimental results so as to obtain our data-driven approach to the problem.

In Fig. 3 we show the result of the embedding of 100 different experiments $\boldsymbol{\Xi} = \{\boldsymbol{x}_1(t_i), \boldsymbol{x}_2(t_i), \ldots, \boldsymbol{x}_{100}(t_i)\}$, $i = 1, \ldots, 5000$, to a manifold in $\mathbb{R}^3$. Each embedded result is plotted according to the $y_0$ value with whom the experimental $x$ values were obtained. It is worth noting that this value is unknown, but $k$-PCA is able to unveil it and, notably, to order the experimental points according to this "hidden" value.

Even if we normally do not have access to the value of the intrinsic variables, like $y_0$, it is extremely important to be able to determine the manifold structure of the data, indicating in this case that a one dimensional arc-length parameter $s$ could be established along the manifold in Fig. 3.

In the following sections we analyze the performance of this technique on two different problems, one discrete, and the second one, continuous.

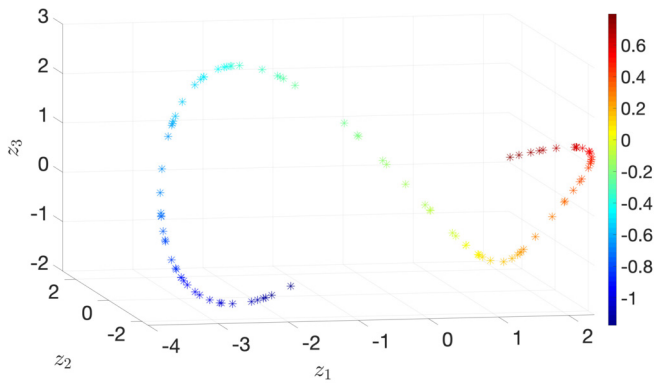



**Fig. 3.** After applying $k$-PCA to the set of 100 experimental results, mapping them back to D=3, we observe that the technique is able to order the results, according to the value of $y_0$, represented in the color bar, even if it is unknown. The axis represents the three first PCA embedding coordinates, whose physical meaning is not known, in general. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

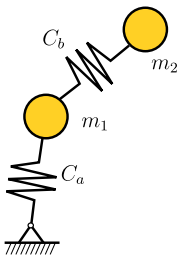



**Fig. 4.** The double pendulum.

## 4. Numerical examples

In this section we analyze two different examples. The first one is discrete, a thermoelastic dissipative pendulum. The second one is continuous, the analysis of a Couette (shear) flow in an Oldroyd-B fluid.

### *4.1. A first example: the double pendulum*

#### *4.1.1. Description of the numerical experiment*

Consider the double pendulum, shown in Fig. 4. This system has a clear Hamiltonian structure. However, let us imagine a situation in which we do not have access to sampling the evolution of the first of the masses, $m_1$. We do not even know anything about its existence. Our current experimental setup only allows us to measure the position and momentum of the second mass, $\boldsymbol{q}_2$ and $\boldsymbol{p}_2$ respectively.

Every experiment considered the same starting position and momentum for the $m_2$ mass—viz., $\boldsymbol{q}_2 = (2.2, 0.0)$ and $\boldsymbol{p}_2 = (1.0, 0.0)$. On the contrary, since we can not control the initial conditions for the mass $m_1$, we assumed a uniform distribution on an interval $\pm 0.02$ around $\boldsymbol{q}_1 = (1.0, 0.0)$ and $\boldsymbol{p}_1 = (0.0, 2.0)$. It is worth noting the small difference of the initial conditions of these experiments, and the chaotic behavior of the pendulum.

In this situation, what we observe is that different measurements on the trajectory of $m_2$ give different results, something that, in the absence of any other information, we interpret as noise, see Fig. 5.

This noise comes, obviously, from the different starting conditions (position $\boldsymbol{q}_1$ and momentum $\boldsymbol{p}_1$) of the first mass for each different execution of the experiment. Since we do not have access to this information, we observe noise, as predicted by the Mori-Zwanzig theory.

#### *4.1.2. Unveiling the existence of internal variables*

Data series $\boldsymbol{\xi}(t) = \{\boldsymbol{q}_2(t), \boldsymbol{p}_2(t)\}$ coming from 80 different experiments are projected, thanks to $k$-PCA, to $\mathbb{R}^8$. Of course, here we leverage our *a priori* knowledge on the true dimensionality of the problem, whose microscopic state is exactly described in $\mathbb{R}^8$. In other words, describing the pendulum by taking $\boldsymbol{z} = \{\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{p}_1, \boldsymbol{p}_2\} \in \mathbb{R}^8$ gives a Hamiltonian structure in which no dissipative character is found. By suppressing any of these variables from the phase portrait, an apparent noise appears (as in Fig. 5) and this fluctuation in the results is equivalent to dissipation—by the well-known fluctuation-dissipation theorem—. In this last case, the addition of a second potential—entropy, see Eq. (1)—is mandatory to describe the system accurately. The result of this projection can not, for obvious reasons, be represented here.

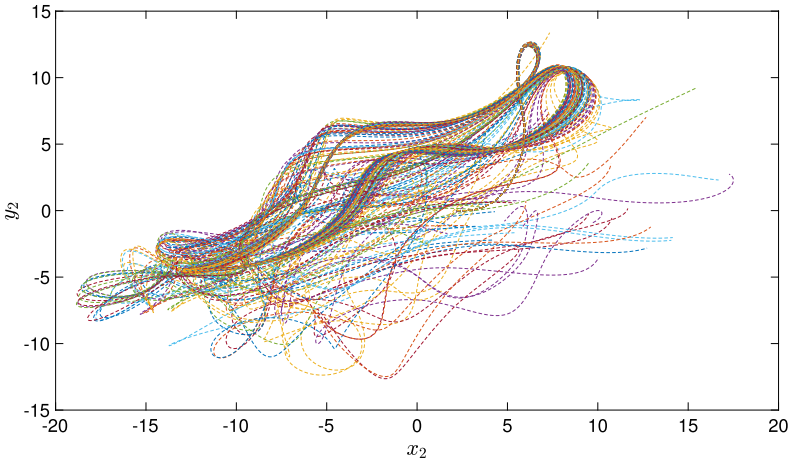



**Fig. 5.** Results for 80 different experiments in which the position of the second mass, $\boldsymbol{q}_2 = (x_2, y_2)$ is measured. Note the high degree of noise in the results.

*4.1.3. Learned behavior*

Once the proposed technique has been able to detect the existence of hidden or internal variables, we are in the position of reconstructing any new experiment.

The GENERIC structure of each one of the 80 different experiments has been obtained by regression, as explained in Eq. (3). The corresponding structure for the new experiment is then interpolated from these 80 results. Of course, this interpolation is made on the data manifold. As in LLE, for instance, we assume a certain number of neighbors for each datum (5 in this case) [42]. For each data point $\boldsymbol{z}^m$ we can write the locally linear data reconstruction as:

$$\boldsymbol{z}^m = \sum_{i \in \mathcal{S}_m} W_{mi} \boldsymbol{z}^i, \tag{6}$$

where $W_{mi}$ are the unknown weights and $\mathcal{S}_m$ the set of the 5-nearest neighbors of $\boldsymbol{z}^m$.

As the same weights appear in different locally linear reconstructions, the best compromise is searched by looking for the weights, all them grouped in vector $\mathbf{W}$, that minimize the functional

$$\mathcal{F}(\mathbf{W}) = \sum_{m=1}^{M} \left\| \boldsymbol{z}^m - \sum_{i=1}^{M} W_{mi} \boldsymbol{z}^i \right\|^2 \tag{7}$$

where here $W_{mi}$ is zero if $\boldsymbol{z}^i$ does not belong to the set of 5-nearest neighbors of $\boldsymbol{z}^m$.

We compare three approaches to reconstruct the new experiment. In the absence of any other information, what experimentalists usually do is to compute the mean of the experimental measurements and fit the model from it. Here, we obtained a *mean* GENERIC model for the results, by assuming $\boldsymbol{\xi} = \{\boldsymbol{q}_2, \boldsymbol{p}_2\} \in \mathbb{R}^4$. This is not, in general a good option, as will readily be noticed. Instead, it is better to find the true neighboring experiments in this space, and to interpolate the GENERIC structure of the experiment at hand from them. We then compared these reconstructions to the one obtained by performing an interpolation of GENERIC models in $\mathbb{R}^6$ and $\mathbb{R}^8$. Results are depicted in Fig. 6.

It can be noticed how the proposed technique provides with predictions several orders of magnitude more accurate than the standard experimental procedures. In fact, the error, measured in 2-norm, is 1.80 % for the proposed technique in $\mathbb{R}^6$, versus 35.22 % for the classical, mean GENERIC one. This error decreases to 1.36% if the interpolation, still in $\mathbb{R}^4$, is performed among the right neighboring experiments in the database. In addition, it is worth noting that the proposed strategy is able to determine that, for a description in $\mathbb{R}^8$, no entropy term in Eq. (1) is necessary, as expected. This means that the resulting manifold is actually equivalent to the microscopic phase space—it has, at least, the same number of dimensions—. In this case, the error in the predicted trajectory resulted to be 1.603%. The ground truth expression for this system has been obtained from [43].

What is noticeable about these results is that, regardless of the scale at which the physics is to be represented, GENERIC provides the right framework for a thermodynamically sound description. If the dissipative part of the equation is found accurately, the physics will be described no matter which scale is employed. This excludes, however, the standard experimental practice, that consists in computing the mean of the experimental results. This practice is demonstrated to be nonsense and should be avoided.

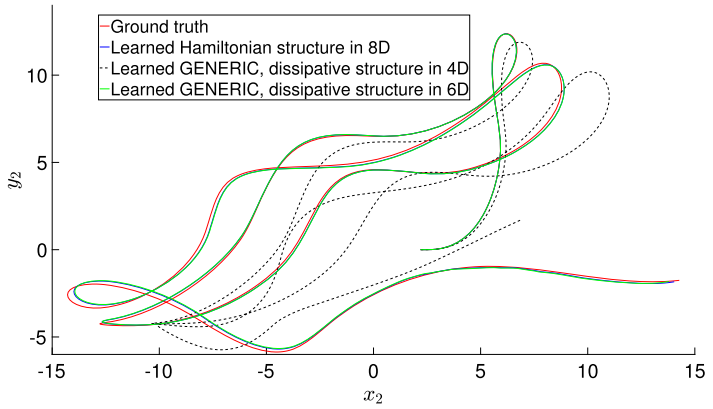



**Fig. 6.** Reconstruction of a new experiment. Comparison among the result predicted by a mean GENERIC model in $\mathbb{R}^4$—dashed black line—the one in $\mathbb{R}^6$, and the one obtained in the appropriate phase space, i.e., $\mathbb{R}^8$—blue line—. The ground truth is represented by the continuous red line.

### *4.2. A second example: viscoelastic fluids*

In continuum mechanics, many models rely on a microscopic representative volume so as to achieve a multiscale description of their constitutive equations. This is actually a way to look for a higher-dimensional constitutive manifold in which the history and noise terms have less influence on the results. Among these models, viscoelastic fluids present many different examples.

To show how the proposed technique works, we have chosen one of the simplest viscoelastic fluid models, the Oldroyd-B model [44]. This a very particular model, that can be obtained by starting from the stochastic description of a suspension of linear elastic dumbbells, or by continuum considerations [45].

#### *4.2.1. The Oldroyd-B model*

The Oldroyd-B model can be seen as arising from considering the deviatoric part $\boldsymbol{T}$ of the stress tensor $\boldsymbol{\sigma}$ (the so-called extra-stress tensor), of the form

$$\boldsymbol{T} + \lambda_1 \overset{\triangledown}{\boldsymbol{T}} = \eta_0 \left( \dot{\boldsymbol{\gamma}} + \lambda_2 \overset{\triangledown}{\dot{\boldsymbol{\gamma}}} \right), \tag{8}$$

where the triangle denotes Oldroyd's upper-convected derivative [46]. Coefficients $\eta_0$, $\lambda_1$ and $\lambda_2$ are material constants. $\dot{\boldsymbol{\gamma}} = (\nabla^s \boldsymbol{v}) = \boldsymbol{D}$ represents the strain rate tensor. $\boldsymbol{v}$ represents the velocity field in the fluid. $\nabla^s \boldsymbol{v}$ stands here for the symmetric part of the gradient operator.

By considering the stress in the solvent (denoted by a subscript $s$) and polymer (denoted by a subscript $p$) components as

$$\boldsymbol{T} = \eta_s \dot{\boldsymbol{\gamma}} + \boldsymbol{\tau},$$

we obtain

$$\boldsymbol{\tau} + \lambda_1 \overset{\triangledown}{\boldsymbol{\tau}} = \eta_p \dot{\boldsymbol{\gamma}},$$

which is the constitutive equation for the elastic stress.

As mentioned before, an alternative derivation can be obtained by starting from a population of linear dumbbells immersed in a Newtonian fluid [44]. In that case, the model relies on the microscopic description of the state of the dumbbells [47]. In this case, a particularly useful choice is to base the microscopic description on the evolution of the conformation tensor $\boldsymbol{c} = \langle \boldsymbol{rr} \rangle$, or, in other words, the second moment of the dumbbell end-to-end distance distribution function. This tensor is in general not experimentally measurable and plays the role of an internal variable.

In this case, the extra-stress tensor results to be proportional to the convective (or Oldroyd) derivative of the conformation tensor,

$$\boldsymbol{\tau} = -\frac{n}{2\zeta_{12}} \overset{\triangledown}{\boldsymbol{c}}.$$

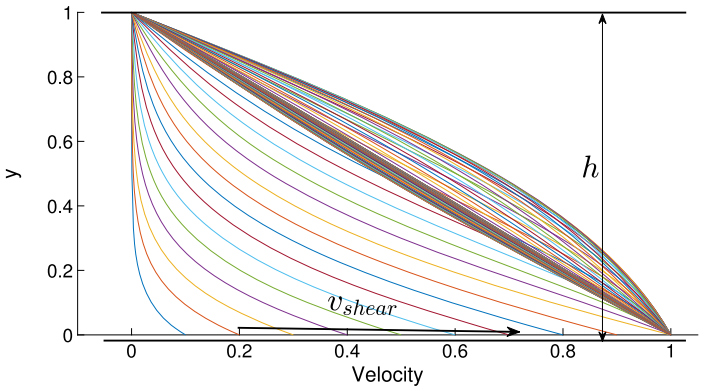



**Fig. 7.** Evolution of the velocity profile at the startup of shear flow in an Oldroyd-B fluid. The different lines represent the velocity profile at different time instants. Note the overshoot in the profile, that tends to the linear profile, typical of a Couette flow, for sufficiently large time. Velocity profile at every $t = 0.01$ seconds ($t \in (0, 1.5]$). Velocity is imposed at the bottom plate, with $h = 1.0$ and $v_{\text{shear}} = 1.0$.

*4.2.2. Model problem*

We consider, as a model problem, the startup of shear (Couette) flow. The evolution of the velocity profile of this flow is shown in Fig. 7. As we did in previous examples, we are going to perform different numerical experiments with different values of the variable $\boldsymbol{c}(t = 0)$ so as to determine under what circumstances the proposed strategy is able to unveil its influence in the model.

The problem is solved by the so-called CONNFFESSIT technique [48]. In it, the probability of finding a dumbbell at a given position in the fluid $\boldsymbol{x}$, at a given time instant $t$, is governed by the corresponding Fokker-Plank equation [45]. Given its inherent high dimensionality, this Fokker-Plank equation is solved advantageously by converting it in its corresponding Itô stochastic differential equation,

$$dr_x(y,t) = \left(\frac{\partial u}{\partial y} r_y(t) - \frac{1}{2\text{We}} r_x(y,t)\right) dt + \frac{1}{\sqrt{\text{We}}} dV_t \tag{9a}$$

$$dr_y(t) = -\frac{1}{2\text{We}} r_y(t) dt + \frac{1}{\sqrt{\text{We}}} dW_t, \tag{9b}$$

where $\boldsymbol{r} = [r_x, r_y]^\top$, and where we have assumed a Couette flow, so that $r_y$ depends solely on $t$, and not on $y$, We is the Weissemberg number and $V_t$, $W_t$ are two independent one-dimensional Brownian motions. This equation is solved by Monte Carlo techniques, by replacing the mathematical expectation by the empirical mean. A total of 10,000 dumbbells were considered at each nodal location in the model, which is composed by 480 nodes along the 200 mm distance between plates, see Fig. 7. In this way, the expected stress ($xy$ component of the stress tensor) will be given by

$$\tau_{xy} = \frac{\epsilon}{\text{We}} \frac{1}{K} \sum_{k=1}^{K} r_x r_y,$$

where, as mentioned before, we have taken $K = 10,000$ dumbbells, and $\epsilon = \frac{\eta_p}{\eta_s}$ represents the ratio of the polymer to solvent viscosities. For the analyzed problem, $\text{Re} = 0.1$ and $\text{We} = 0.1$ numbers were considered.

*4.2.3. Learning the model: unveiling the origin of noise*

CONNFFESSIT simulations usually take isotropic dumbbell distributions to begin with. Here, however, we have performed up to 750 simulations by starting from different conformation (and therefore also stress) tensors. Once subjected to the analysis of $k$-PCA techniques, the resulting eigenvalues of the data are shown in Fig. 8. Note in the detail of the figure that the three first eigenvalues, whose value descends rapidly, are one order of magnitude bigger than the following ones. This suggests the presence of three to six dimensions more important than the others. Beyond the sixth eigenvalue, they descend in a less pronounced way. This reasoning is of course qualitative, and no definitive response exists for the estimation of the intrinsic dimension of a manifold. There seems to be no other alternative to perform some numerical experiments and to check the goodness of the results.

These differences in eigenvalues may not be so pronounced as in other cases, but remember that, due to the Monte Carlo approach to the problem, the stress tensor becomes in fact a stochastic variable, thus equipped with some unavoidable noise.

We embed these experiments onto the space spanned by the first four eigenvectors of the $k$-PCA. In Fig. 9 we plot the mentioned 750 experiments on a three-dimensional plot, where a sphere-like manifold structure is observed. This also suggests the validity of an embedding onto a three-dimensional manifold.

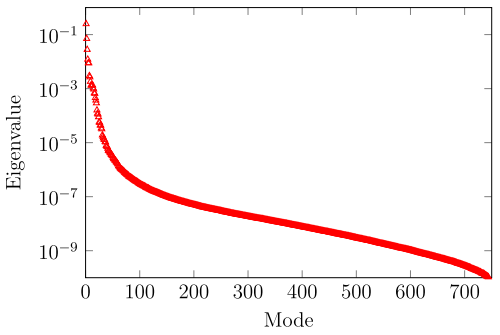



**Fig. 8.** Eigenvalues of the sample covariance matrix for the Oldroyd-B problem.

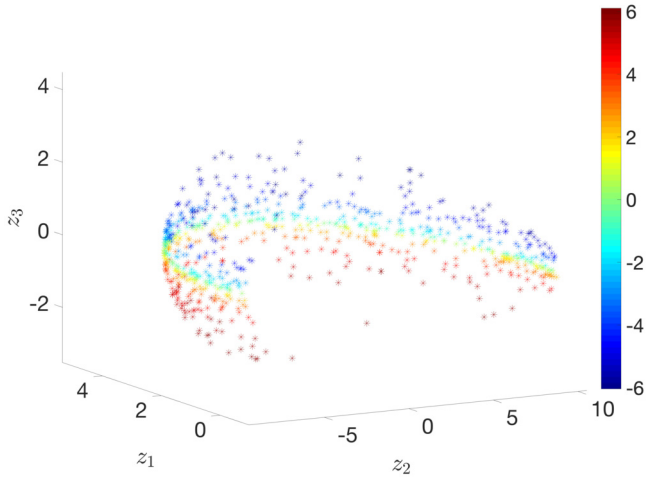



**Fig. 9.** Embedding of 750 Oldroyd-B results onto the space spanned by the first four eigenvalues of the $k$-PCA. For obvious reasons, it is not possible to plot them in a three-dimensional plot. Here, the first three coordinates of the resulting embedding have been conserved. This also suggests that an embedding onto a three-dimensional manifold would have provided similar results.

The legend (and the color in the dots representing each experiment) in Fig. 9 represents the value of the stress component $\tau_{xy}$ at the startup. We notice that the results are very well clustered according to this value, even if very similar results are obtained by plotting the $xy$ component of the conformation tensor, for instance or, less clearly, the value $\arctan(r_x/r_y)$ at $t = 0$, for instance.

This fact suggests that, even if we can not measure it, the proposed technique is able to infer that results depend on a variable other than the macroscopic ones. This variable seems to live in a three- or four-dimensional manifold (up to six dimensions have been considered with no apparent gain in accuracy). The embedding in three dimensions shows good results (the fourth dimension could be understood as the inherent noise in the simulation). These three dimensions are interpreted by us as equivalent to the three distinct components of the conformation tensor or the three distinct components of the stress tensor.

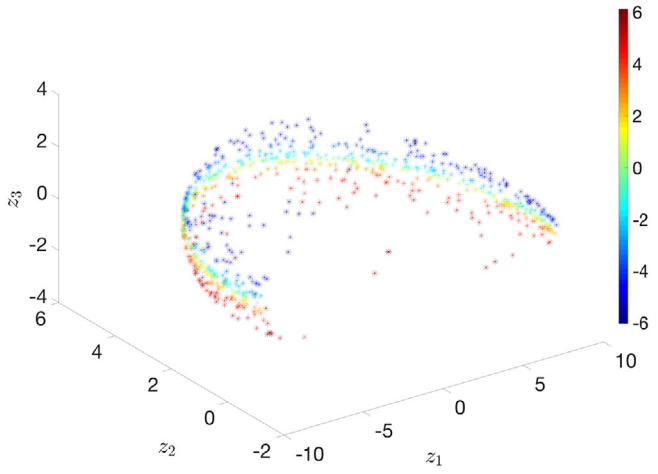
For each one of these 750 simulations, a GENERIC model has been obtained by regression, see Section 2. This constitutes the learning phase of our method, in which we obtain a thermodynamic consistent representation of these flows.

*4.2.4. Learning the model: unveiling the dependence on history*

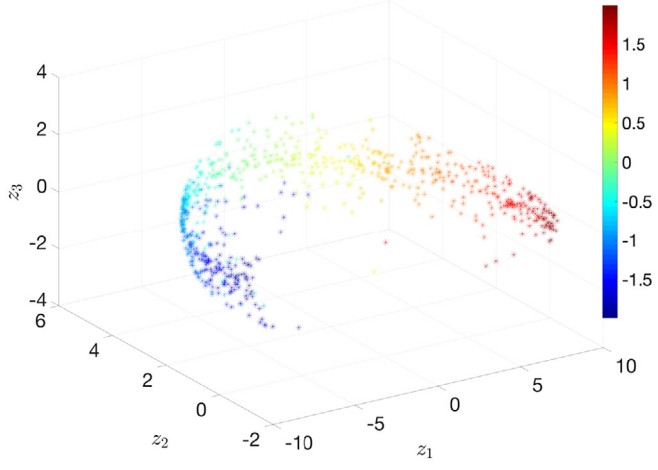
The strain rate at which the experiment is performed is something that we can control completely, to a great accuracy. And, of course, it will affect the results by changing the history of the experiment—it plays the same role as the non-Markovian term in Eq. (5)—. Up to 750 different experiments were performed in which both the initial configuration of the dumbbells and the shear rate were considered as random variables. The goal was to determine if the proposed method is able to ascertain the influence of both, history and noise, in the experimental results.

As can be noticed from Fig. 10, results are now distributed on a manifold that resembles the shape of the outer surface of a torus or ellipsoid. Moreover, if we plot the embeddings attending to the initial state of the dumbbells or the shear rate, respectively, we obtain that the experiments are correctly clustered according to these values.

Note that, in a general situation, we will not have any information about the nature of the found internal variables. Nevertheless, the suggested method still provides with valuable information on the number of dimensions in which the problem should be described, even in the absence of any physical interpretation about the meaning of these coordinates.




**(a)** Embedding according to the initial dumbbell orientation. The legend reflects the value of the $\tau_{xy}$ stress tensor component.




**(b)** Embedding according to the value of shear rate. In this case, the legend reflects the value of the shear rate, $\dot{\gamma}$. during the experiment.

**Fig. 10.** Result of the embedding of 750 different experiments according to (a) the noise due to the initial state of the dumbbell distribution and (b) the value of the shear rate. Observe how in both cases the proposed method is able to unveil the existence of the internal, hidden variables and, noteworthy, to suggest an embedding on a manifold of the expected dimensionality. Both color plots suggest that the topology of the resulting manifold could be parameterized by two coordinates, one in meridian and the other in parallel directions. These coordinates, whose physical meaning is not provided by the method, are identified by us as the conformation tensor and the shear rate.

#### *4.2.5. Predictive capabilities of the method*

But, of course, the goal of scientific learning is to be able to make predictions about the future. Therefore, nothing in the above procedure would be useful if we are not able to predict the evolution of an arbitrary Oldroyd-B flow. The predictive capabilities of the method will thus be given by its ability to "learn" the GENERIC structure of an arbitrary flow, not considered in the database.

*Characterizing noise* To that end, we consider a new example, that starts from a new configuration tensor, not considered before, in the first experimental set—that obtained with constant shear rate—. This new flow is denoted with a big black dot in Fig. 9. This allows us, in fact, to determine which experimental results are "neighbors" of this new situation in the space dictated by the *k*-PCA projection. In other words: which of the already learned examples are the closest to this new one in terms of the hidden, three-dimensional internal variable just found.

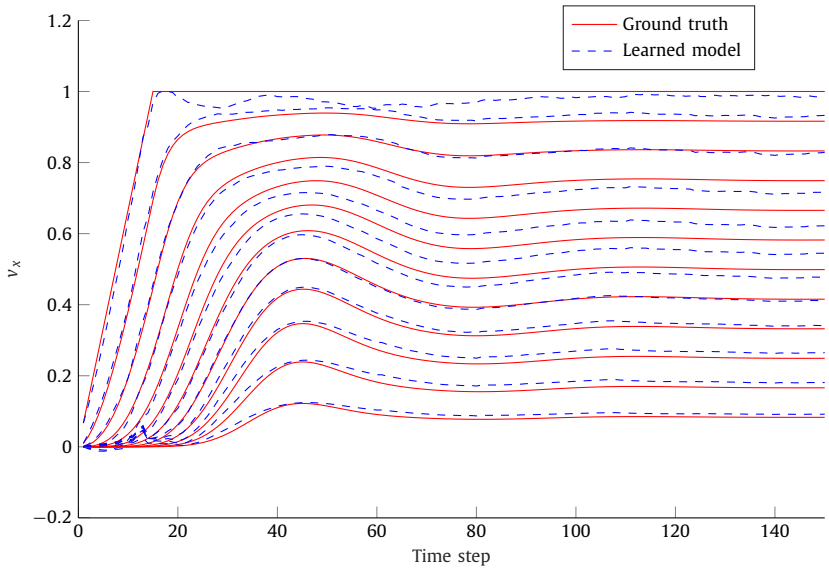



**Fig. 11.** Evolution in time of the velocity $v_x$ for twelve distinct nodes in the mesh. The continuous red line represents the ground truth, obtained by direct numerical solution, while the dashed line represents the learned GENERIC model for the flow.

We employ again the strategy introduced in Section 4.1.3, by choosing $M = 3$ neighbors. This value is of course a user parameter, as in many other non-linear dimensionality reduction techniques [42]. The new GENERIC model for the just considered flow will therefore be given by locally linear interpolation of the neighboring GENERIC models, thus obtaining a totally predictive and thermodynamically sound model for this problem.

It is worth noting that this new model takes into consideration the resolved variables considered in the original GENERIC models for the examples in the experimental database, but also the dependence of the model on this new, three-dimensional hidden or unresolved variables. So to speak, we have increased the dimensionality of the description by learning the presence of an internal variable.

The evolution in time of the velocity component $v_x$ for twelve nodes out of the 480 of the mesh is shown in Fig. 11. The error in discrete 2-norm was found to be between 2.24 and 4.3%, depending on the particular conditions of the execution.

*Effect of history* If we repeat the same strategy for the big dataset in which we make vary both the initial condition (i.e., noise) and shear rate (thus affecting history), we obtain similar accuracy levels. In this case, the dimension of the embedding space was augmented, and assumed to be six. This is due, of course, to the expected increase on the true dimensionality of the data, motivated by the dependence on the shear rate.

By taking again three neighbors for each experiment, a randomized new experiment is reproduced with an error of 3.77% in discrete 2-norm.

## 5. Conclusions

While data-driven learning of physical laws possesses an inherent interest for obvious reasons, it has lacked of a general theory about the right scale at which the data must be represented or analyzed. In sum, at what scale does the resulting law should be written. It is well known that a coarse-grained description of the results gives rise to the appearance of history-dependent and noisy laws, while a very microscopic, detailed representation leads to an overwhelming amount of information to be processed.

The method here presented constitutes the first attempt—to the best of our knowledge—to unveil the need of internal variables in the description of a physical phenomenon. By employing $k$-PCA techniques, we have shown that data informs us about the right dimensionality in which these results should be represented in the form of a physical law.

But experiments usually proceed in an autonomous manner, by sampling at scales dictated by the current technological knowledge and not by the physics itself. If this is the case, the method here proposed suggests the employ of internal variables and also indicates the number, or at least an approximation to it, of these variables.

The main limitation of this approach, however, is that it does not provide with any information on the physical nature of these internal variables. The reasoning of well-educated physicists remains to be indispensable to unveil the true nature of these variables.

This does not constitute a problem, however, for many applications of industrial interest. For instance, this technique opens the possibility of constructing digital twins able to learn by themselves, just by analyzing data, and to correct, if

necessary, not only the parameter values of their implemented models, but the models themselves. This constitutes part of our current effort of research.

## CRediT authorship contribution statement

David González: software, investigation, validation.
Francisco Chinesta: methodology, validation.
Elías Cueto: conceptualization, methodology, investigation, writing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References


[1] Steven L. Brunton, Joshua L. Proctor, J. Nathan Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, Proc. Natl. Acad. Sci. (2016).
[2] Kadierdan Kaheman, J. Nathan Kutz, Steven L. Brunton. Sindy-pi, A robust algorithm for parallel implicit sparse identification of nonlinear dynamics, arXiv preprint arXiv:2004.02322, 2020.
[3] T. Kirchdoerfer, M. Ortiz, Data-driven computational mechanics, Comput. Methods Appl. Mech. Eng. 304 (2016) 81–101.
[4] Jacobo Ayensa-Jiménez, Mohamed H. Doweidar, Jose A. Sanz-Herrera, Manuel Doblaré, A new reliability-based data-driven approach for noisy experimental data with physical constraints, Comput. Methods Appl. Mech. Eng. 328 (2018) 752–774.
[5] M.A. Bessa, R. Bostanabad, Z. Liu, A. Hu, Daniel W. Apley, C. Brinson, Wei Chen, Wing Kam Liu, A framework for data-driven analysis of materials under uncertainty: countering the curse of dimensionality, Comput. Methods Appl. Mech. Eng. 320 (2017) 633–667.
[6] Julian Lißner, Felix Fritzen, Data-driven microstructure property relations, Math. Comput. Appl. 24 (2) (2019) 57.
[7] R. Ibañez, E. Abisset-Chavanne, D. Gonzalez, J.L. Duval, E. Cueto, F. Chinesta, Hybrid constitutive modeling: data-driven learning of corrections to plasticity models, Int. J. Mater. Forming 12 (2019) 717–725.
[8] Marcos Latorre, Francisco Javier Montáns, What-you-prescribe-is-what-you-get orthotropic hyperelasticity, Comput. Mech. 53 (6) (Jun 2014) 1279–1298.
[9] Ruben Ibañez, Domenico Borzacchiello, Jose Vicente Aguado, Emmanuelle Abisset-Chavanne, Elias Cueto, Pierre Ladeveze, Francisco Chinesta, Data-driven non-linear elasticity: constitutive manifold construction and problem discretization, Comput. Mech. 60 (5) (Nov 2017) 813–826.
[10] Rubén Ibanez, Emmanuelle Abisset-Chavanne, Jose Vicente Aguado, David Gonzalez, Elias Cueto, Francisco Chinesta, A manifold learning approach to data-driven computational elasticity and inelasticity, Arch. Comput. Methods Eng. 25 (1) (2018) 47–57.
[11] E. Lopez, D. Gonzalez, J.V. Aguado, E. Abisset-Chavanne, E. Cueto, C. Binetruy, F. Chinesta, A manifold learning approach for integrated computational materials engineering, Arch. Comput. Methods Eng. (2016) 1–10.
[12] Maziar Raissi, Paris Perdikaris, George Em Karniadakis, Physics informed deep learning (part I): data-driven solutions of nonlinear partial differential equations, arXiv preprint, arXiv:1711.10561, 2017.
[13] Maziar Raissi, Paris Perdikaris, George Em Karniadakis, Physics informed deep learning (part II): data-driven discovery of nonlinear partial differential equations, arXiv preprint, arXiv:1711.10566, 2017.
[14] Maziar Raissi, Alireza Yazdani, George Em Karniadakis, Hidden fluid mechanics: learning velocity and pressure fields from flow visualizations, Science 367 (6481) (2020) 1026–1030.
[15] D. González, F. Chinesta, E. Cueto, Thermodynamically consistent data-driven computational mechanics, Contin. Mech. Thermodyn. 31 (2019) 239–253.
[16] David Gonzalez, FRancisco Chinesta, Elías Cueto, Learning corrections for hyperelastic models from data, Front. Mater. 6 (14) (2019).
[17] Miroslav Grmela, Hans Christian Öttinger, Dynamics and thermodynamics of complex fluids. I. Development of a general formalism, Phys. Rev. E 56 (Dec 1997) 6620–6632.
[18] M. Pavelka, V. Klika, M. Grmela, Multiscale Thermodynamics, De Gruyter, 2018.
[19] Francisco Chinesta, Elias Cueto, Miroslav Grmela, Beatriz Moya, Michal Pavelka, Learning physics from data: a thermodynamic interpretation, arXiv preprint arXiv:1909.01074, 2019.
[20] Robert Eggersmann, Trenton Kirchdoerfer, Stefanie Reese, Laurent Stainier, Michael Ortiz, Model-free data-driven inelasticity, Comput. Methods Appl. Mech. Eng. (2019).
[21] Norbert Schaudinnus, Andrzej J. Rzepiela, Rainer Hegger, Gerhard Stock, Data driven Langevin modeling of biomolecular dynamics, J. Chem. Phys. 138 (20) (2013) 204106.
[22] Huan Lei, Nathan A. Baker, Xiantao Li, Data-driven parameterization of the generalized Langevin equation, Proc. Natl. Acad. Sci. 113 (50) (2016) 14183–14188.
[23] Illia Horenko, Carsten Hartmann, Christof Schütte, Frank Noe, Data-based parameter estimation of generalized multidimensional Langevin processes, Phys. Rev. E 76 (1) (2007) 016706.
[24] Norbert Schaudinnus, Andrzej J. Rzepiela, Rainer Hegger, Gerhard Stock, Data driven Langevin modeling of biomolecular dynamics, J. Chem. Phys. 138 (20) (2013) 204106.
[25] Rich Friedrich, Silke Siegert, J. Peinke, St. Lueck, Marcus Siefert, Michael Lindemann, Jan Raethjen, Güntner Deuschl, Gerhard Pfister, Extracting Model Equations from Experimental Data, 2000.
[26] Hazime Mori Transport, Collective motion, and Brownian motion, Prog. Theor. Phys. 33 (3) (1965) 423–455.
[27] Robert Zwanzig, Nonlinear generalized Langevin equations, J. Stat. Phys. 9 (3) (1973) 215–220.
[28] Yuanran Zhu, Daniele Venturi, Generalized Langevin equations for systems with local interactions, J. Stat. Phys. (2020) 1–31.
[29] Yuanran Zhu, Daniele Venturi, Faber approximation of the Mori–Zwanzig equation, J. Comput. Phys. 372 (2018) 694–718.
[30] Yuanran Zhu, Jason M. Dominy, Daniele Venturi, On the estimation of the Mori-Zwanzig memory integral, J. Math. Phys. 59 (10) (2018) 103501.
[31] Catherine Brennan, Daniele Venturi, Data-driven closures for stochastic dynamical systems, J. Comput. Phys. 372 (2018) 281–298.
[32] Daniele Venturi, George Em Karniadakis, Convolutionless Nakajima–Zwanzig equations for stochastic analysis in nonlinear dynamical systems, Proc. R. Soc. A, Math. Phys. Eng. Sci. 470 (2166) (2014) 20130754.
[33] E. Weinan, A proposal on machine learning via dynamical systems, Commun. Math. Stat. 5 (1) (Mar 2017) 1–11.
[34] F. Regazzoni, L. Dedè, A. Quarteroni, Machine learning for fast and reliable solution of time-dependent differential equations, J. Comput. Phys. 397 (2019) 108852.

[35] H. Nyquist, Thermal agitation of electric charge in conductors, Phys. Rev. 32 (Jul 1928) 110–113.
[36] Zhong Yi Wan, Pantelis Vlachas, Petros Koumoutsakos, Themistoklis Sapsis, Data-assisted reduced-order modeling of extreme events in complex dynamical systems, PLoS ONE 13 (5) (2018) 1–22.
[37] Pep Español, Statistical Mechanics of Coarse-Graining, Springer, Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 69–115.
[38] Chao Ma, Jianchun Wang, E. Weinan, Model reduction with memory and the machine learning of dynamical systems, arXiv:1808.04258 [abs], 2018.
[39] Bernhard Schölkopf, Alexander Smola, Klaus-Robert Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (5) (July 1998) 1299–1319.
[40] B. Scholkopf, A. Smola, K.R. Muller, Kernel principal component analysis, in: Advances in Kernel Methods - Support Vector Learning, MIT Press, 1999, pp. 327–352.
[41] J.A. Lee, M. Verleysen, Nonlinear Dimensionality Reduction, Springer Verlag, 2007.
[42] Sam T. Roweis, Lawrence K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[43] Ignacio Romero, Thermodynamically consistent time-stepping algorithms for non-linear thermomechanical systems, Int. J. Numer. Methods Eng. 79 (6) (2009) 706–732.
[44] R.G. Owens, T.N. Phillips, Computational Rheology, Imperial College Press, 2002.
[45] Claude Le Bris, Tony Lelièvre, Multiscale modelling of complex fluids: a mathematical initiation, Technical Report RR-627, INRIA, 2007.
[46] K. Walters, M.F. Webster, The distinctive CFD challenges of computational rheology, Int. J. Numer. Methods Fluids 43 (5) (2003) 577–596.
[47] Matteo Pasquali, L.E. Scriven, Theoretical modeling of microstructured liquids: a simple thermodynamic approach, J. Non-Newton. Fluid Mech. 120 (1) (2004) 101–135, 3rd International Workshop on Nonequilibrium Thermodynamics and Complex Fluids.
[48] Manuel Laso, Hans Christian Öttinger, Calculation of viscoelastic flow using molecular models: the connffessit approach, J. Non-Newton. Fluid Mech. 47 (1993) 1–20.